Faculty of Engineering and Information Technology
University of Technology Sydney

Friend Recommendation in Social Multimedia Networks

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Shangrong Huang

June 2017

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by myself. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Acknowledgments

Foremost, I would like to express my sincere gratitude to my supervisors Associate Prof. Jian Zhang for continued support of my PhD study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance has helped me in all aspects of research and writing this thesis. I could not have imagined having a better advisor and mentor for my PhD study.

I also would like to appreciate Associate Prof. Qiang Wu for providing me with continuous support throughout my PhD study and research.

Associate Prof. Lei Wang at University of Wollongong, Dr. Xiansheng Hua from Alibaba Group China, Dr. Tao Mei from Microsoft Asia, Dr. Shiyang Lu at NICTA Australia, and Dr. Xinwang Liu have also given me a lot of academic suggestions and supports. Without their professional guidance and persistent help, this thesis would not have been possible.

I thank my fellows in Global Big Data Technologies Centre: Jingsong Xu, Yucheng Wang, Hao Cheng for the stimulating discussions, and for all the fun we have had in the last several years.

I thank Dr. Jing Ren for her support during my PhD study.

Last but not the least, I would like to thank my parents, for their unconditional support, both financially and emotionally throughout my entire PhD journey.

Shangrong Huang
June 2017 @ UTS

Contents

List of Figures

List of Tables

List of Publications

Papers Published

- **Shangrong Huang**, Jian Zhang, Dan Schonfeld, Lei Wang, and Xian-Sheng Hua, Two-Stage Friend Recommendation Based on Network Alignment and Series-Expansion of Probabilistic Topic Model, *IEEE Transactions on Mutlimedia*, Volumn 19, Issue 6, pp.1314-1326.

- **Shangrong Huang**, Jian Zhang, Lei Wang, and Xian-Sheng Hua (2016), Social Friend Recommendation Based on Multiple Network Correlation. *IEEE Transactions on Mutltimedia*, Volumn 18, Issue 2, pp. 287-299.

- **Shangrong Huang**, Jian Zhang, Shiyang Lu, and Xian-Sheng Hua (2015), Social Friend Recommendation Based on Network Correlation and Feature Co-Clustering. *in Proceedings of the 5th ACM International Conference on Multimedia Retrieval* (**ICMR'15**) , pp. 315-322.

- **Shangrong Huang**, Jian Zhang, Xinwang Liu, and Lei Wang (2014), A Method of Discriminative Information Preservation and In-Dimension Distance Minimization for Feature Selection. *in Proceedings of the International Conference on Pattern Recognition* (**ICPR'14**) , pp. 1615-1620

Abstract

With the rapid development of computer science and internet technologies, social media and social network has experienced explosive growth over the last decades. Social websites, such as Flickr, YouTube, and Twitter, have billions of users who share photos, videos and opinions, they also make friends on these websites. On-line friendship is an emerging topic that attracts the attentions from both economists and sociologists. The study of the on-line friendship, on one hand, can help the on-line merchants to find their potential customers, and thus make more precise recommendations; on the other hand, it helps to get a deep understanding of the relationships among different people. However, individuals' on-line friend making behaviour is relatively complex and may be affected by many different factors. For example, an individual might make on-line friends with others because they discuss a hard mathematical problem, or it is possible that he/she makes a friend because they both enjoy a film. The reasons for friend making behaviours are likely to be diverse. Traditional friend recommendations that have been widely applied by Facebook and Twitter are often based on common friends and similar profiles such as having the same hobbies or working on a similar topic, which usually can not make a precise recommendation, due to the complexity of the problem. In this thesis, I, with my collaborators, try to give some solutions of on-line social friend recommendation from several aspects. In general, I contribute more than 85% of this thesis.

One problem for social friend recommendation is that how shall we find the important social features that would highly influence individuals' friend

making behaviours. Usually, the reason an individual A would make friends with another person B is not that A is satisfied with all the characteristics of B, but that he/she has interest in some factors that B has illustrated. These factors can be viewed as instructive social features for friend recommendation tasks. So in this thesis, we first discuss the important social features for friend recommendation.

Chapter 3 provides a general algorithm of important feature selection that can be applied in different fields such as biological and face image classification. The idea is to project the high dimensional data into lower dimensional space and select the important features that preserve both the global and local similarity structures of the datasets.

Chapter 4 extends the basic idea of Chapter 3 to the field of social networks, and consider the friend recommendation task from the view of the network structure. First we consider the tag features. The important tag features are chosen so that the Flickr tag similarity network looks similar to the Flickr contact network. In other words, Flickr tag similarity network is aligned to the contact network by selecting the important tag features. This network alignment method can also be applied to more than one networks.

In Chapter 5 we begin to take the image features into consideration. It would be relatively difficult to analyse the multi-domain data simultaneously. In this thesis we design a multi-stage scenario to consider the information from one domain in one stage. In this way, not only the complexity of the problem is reduced, but we can also make a deep analysis about the contributions of the information from different domains. For the algorithm proposed in Chapter 5, for the first stage we utilise the tag information similarly as the method suggested in Chapter 4, for the second stage we propose a co-clustering method that clusters the contact information, tag and image feature information simultaneously to refine the final recommendation result.

To further improve the recommendation accuracy, in Chapter 6 we apply a topic model based method in the second stage, instead of the co-clustering

method proposed in Chapter 5. The reason for the improvement is that co-clustering method can not provide a precise rank of the recommendation list, but the topic model can give a quantitative analysis of the friendship between two individuals. In this chapter we also provide a new method to find the solution of the topic model, which is different from the widely applied Gibbs sampling, variational inference or the matrix factorization method. The idea is to analytically express the solution of the integral of two random variables, in a series form. In this way we can determine the solution of the probabilistic model precisely, which is better than the traditional Gibbs sampling, variational inference or matrix factorization methods.

In Chapter 7, with the help of widely discussed Deep Learning (DL) Framework, we develop a staged DL-based friend recommendation method. In the first stage, the text and image information is correlated to learn some features via convlutional neural network. In the second stage, the features are refined by the users' clustering information via another deep neural network.

The methods mentioned in Chapter 4, 5, 6 and 7 are applied in a dataset that collected from the widely used image sharing website Flickr. It contains tens of thousands of users, hundreds of thousands tags and millions of images to predict the on-line friendship between users. The performance of these recommendation methods is examined by precision, recall and F-measure. These methods give some insightful knowledge about individuals' online relationship and we hope these methods can help social websites to design their recommendation algorithms.